Geometric Conditions for Subspace-Sparse Recovery

Chong You and René Vidal
Center for Imaging Science, Johns Hopkins University, Baltimore, MD, 21218, USA

***Abstract*—Given a dictionary Π and a signal $\xi = \Pi\mathbf{x}$ generated by a few *linearly independent* columns of Π, classical sparse recovery theory deals with the problem of uniquely recovering the sparse representation $\mathbf{x}$. In this work, we consider the more general case where ξ lies in a low-dimensional subspace spanned by a few possibly *linearly dependent* columns of Π. In this case, $\mathbf{x}$ may not unique, and the goal is to recover any subset of the columns of Π that spans the subspace containing ξ. We call such a representation $\mathbf{x}$ *subspace-sparse*. We study conditions under which existing pursuit methods recover a subspace-sparse representation. Such conditions reveal important geometric insights and have implications for classical sparse recovery as well as subspace clustering.**

I. Subspace-sparse Recovery Problem

Given a dictionary Π and a signal $\xi := \Pi\mathbf{x}$ spanned by M columns of Π, the task of sparse recovery is to find the M-sparse vector $\mathbf{x}$. Let Φ be the matrix containing the M columns that generate ξ. For this problem to be well posed, Φ needs to have full column rank, i.e., $s = M$, where $s := \text{rank}(\Phi)$. If $s < M$, the columns of Φ lie in a low-dimensional subspace and there could be multiple representations of ξ. This issue is faced in the subspace classification [1] and subspace clustering [2] problems, where it is sufficient to identify any matrix Φ that spans the subspace (class or group) containing the signal ξ. We call the corresponding representation $\mathbf{x}$, whose nonzero entries correspond to a subset of the columns of Φ, as *subspace-sparse*.

II. Subspace-Sparse Recovery Conditions

This paper studies conditions under which a subspace-sparse representation can be found using existing sparse recovery methods, such as Orthogonal Matching Pursuit (OMP) or Basis Pursuit (BP). Prior work [2], [3] has derived such conditions in the context of subspace clustering, where ξ is one of the columns of Φ. Here, we study the more general case where ξ is *any* point in the span of Φ.

Before proceeding further, we will need some definitions to characterize Φ. Let $\mathcal{K}(\pm\Phi) = \text{conv}\{\pm\phi_1, \cdots, \pm\phi_s\}$ be the convex hull of the symmetrized columns of Φ, which we call *inlier points*, and let $\mathcal{K}^o(\pm\Phi) = \{\eta \in \mathcal{R}(\Phi) : \|\Phi^\top\eta\|_\infty \le 1\}$ be its polar set, where $\mathcal{R}(\cdot)$ is the range of a matrix. To characterize the "distribution" of the inliers, we define the *inradius* of $\mathcal{K}(\pm\Phi)$, denoted by $r(\mathcal{K}(\pm\Phi))$, as the radius of the largest Euclidean ball in $\mathcal{R}(\Phi)$ that is inscribed in $\mathcal{K}(\pm\Phi)$. Notice that the inradius is relatively large if the inliers are distributed across the entire subspace $\mathcal{R}(\Phi)$ and not skewed towards a certain direction. We will also use the following definition of the "dual" of the inlier points.

Definition 1. *The set of dual points of the matrix Φ, denoted by $\mathcal{D}(\Phi)$, is defined as the set of extreme points of the set $\mathcal{K}^o(\pm\Phi)$.*

An important fact about $\mathcal{D}(\Phi)$ is that it is a finite subset of $\mathcal{R}(\Phi)$. See Figure 1 for an illustration of the definition.

We are now ready to state the first condition for subspace-sparse recovery. Our results assume that the columns of the dictionary Π are of unit norm.

Theorem 1. *Assume that the dictionary $\Pi = [\Phi, \Psi]$ satisfies the following Principal Recovery Condition (PRC):*

$$r(\mathcal{K}(\pm\Phi)) > \max_{\eta \in \mathcal{R}(\Phi)\setminus\{0\}} \|\Psi^\top\eta\|_\infty / \|\eta\|_2. \tag{1}$$

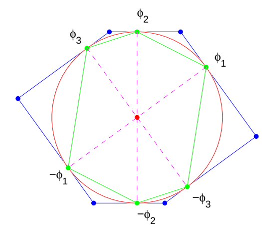

Fig. 1. Illustration of inlier characterizations. Assume $\Phi = [\phi_1, \phi_2, \phi_3]$ spans a two-dimensional subspace. $\mathcal{K}(\pm\Phi)$, $\mathcal{K}^o(\pm\Phi)$ and $\mathcal{D}(\Phi)$ are denoted by the green polygon, the blue polygon, and the six blue dots, respectively.

Then BP and OMP give subspace-sparse solutions for any $\xi \in \mathcal{R}(\Phi)$.

Notice that the LHS of the PRC requires that the inliers be evenly distributed across $\mathcal{R}(\Phi)$ so that $r(\mathcal{K}(\pm\Phi))$ is large. The RHS requires that any column of Ψ and any points in $\mathcal{R}(\Phi)$ should not be too close.

Our second result improves over PRC by showing that it is enough to have a *finite* subset of $\mathcal{R}(\Phi)$ be not too close to columns of Ψ.

Theorem 2. *Assume that the dictionary $\Pi = [\Phi, \Psi]$ satisfies the following Dual Recovery Condition (DRC):*

$$r(\mathcal{K}(\pm\Phi)) > \max_{\eta \in \mathcal{D}(\Phi)\setminus\{0\}} \|\Psi^\top\eta\|_\infty / \|\eta\|_2. \tag{2}$$

Then BP and OMP give subspace-sparse solutions for any $\xi \in \mathcal{R}(\Phi)$.

Thus, by using DRC, Theorem 2 gives a stronger result.

We can apply PRC/DRC also to sparse recovery and the following result can be established. If, for any division of Π into $\Phi \in \mathbb{R}^{n\times s}$ and Ψ, it has $\text{rank}(\Phi) = s$ and PRC/DRC holds, then sparse recovery of any s-sparse vector can be achieved by OMP and BP. Another well known sufficient condition for this purpose is $\mu(\Pi) < \frac{1}{2s-1}$ [4], where $\mu(\Pi)$, the coherence, is the largest absolute inner product between columns of Π. They compare as follows.

Theorem 3. *Given a dictionary Π. If it has $\mu(\Pi) < \frac{1}{2s-1}$, then PRC holds for Π with any partition $\Pi = [\Phi, \Psi]$ where $\Phi \in \mathbb{R}^{n\times s}$.*

This shows that the condition of PRC is a better characterization of sparse recovery than using the coherence parameter.

Acknowledgments: The authors thank the support of NSF BIGDATA grant 1447822.

References

[1] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[2] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[3] M. Soltanolkotabi and E. J. Candès, "A geometric analysis of subspace clustering with outliers," *Annals of Statistics*, vol. 40, no. 4, pp. 2195–2238, 2012.

[4] J. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.